Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 6196810, 12 pages
https://doi.org/10.1155/2022/6196810

Research Article

Foreground Information-Aware Image Superresolution Reconstruction for Image Processing IoT Systems in Smart City

Yanfen Cheng,[1] **Chenhao Li**,[1] **Xun Shao**,[2] **and Fan He**[1]

[1]*School of Computer Science and Artificial Intelligence, Wuhan University of Technology, Wuhan 430063, China*
[2]*School of Regional Innovation and Social Design Engineering, Kitami Institute of Technology, Kitami, Japan*

Correspondence should be addressed to Xun Shao; x-shao@ieee.org and Fan He; hefan@whut.edu.cn

Received 15 November 2021; Revised 15 December 2021; Accepted 21 December 2021; Published 18 January 2022

Academic Editor: Han Liu

In recent years, with the rise of Internet of Things (IoT), a majority of smart technologies, such as autonomous vehicles, smart healthcare, and urban surveillance, require a huge number of images of high quality and resolution. Currently, image superresolution reconstruction technologies are widely used for obtaining high quality images. Unfortunately, the existing methods generally focus on the whole image without highlighting foreground information and lack visual focus. Also, they have low utilization of shallow features and numerous training parameters. In this paper, we propose a feature extraction module that focuses on foreground information: the parallel attention module (PAM). PAM computes channel and spatial attention in parallel, inputs the obtained attention values into a cascaded gated network, and dynamically adjusts the weights of both using nonuniform joint loss to focus on image foreground information and detail features to improve the reconstructed image's foreground sharpness. To further improve the performance, we propose to connect multiple PAM modules in series with skip connections and call it PAMNet. PAMNet can better leverage the shallow residual features, and the reconstructed images are closer to ground truth. Thereby, the applications in the urban image processing IoT systems can obtain high-resolution images more quickly and precisely. The comprehensive experimental results show that PAMNet performs better than the state-of-the-art technologies.

1. Introduction

With the rapid development of artificial intelligence (AI) [1–5] and 5G [6, 7], many emerging technologies, such as Internet of Things [8–17], blockchain [18–21] autonomous vehicles [22–24], smart healthcare [25–31], and urban surveillance, that meet people's aspirations for a better life, are developing very fast. In these smart technologies, image processing IoT applications such as autonomous vehicles, smart healthcare, and urban surveillance are playing important roles in the upcoming smart society. Figure 1 shows the application of urban IoT systems.

However, due to the heterogenous properties of the smart camera devices and complicated network environment, the smart applications deployed in the remote cloud can often only obtain low-resolution images, which largely limit the usage of the smart applications. For example, (1) high-speed cars need to recognize the contents of road signs as early as possible, but due to the long shooting distance and small road signs, it is necessary to convert the captured low-resolution images into high-resolution images with the help of image superresolution methods. (2) In suburban community hospitals, we need to superresolve the transmitted low-resolution images to improve the accuracy of doctors' remote diagnosis due to the poor quality of the captured equipment. (3) Police often tracks the trajectory of suspects through urban surveillance systems, and after image superresolution reconstruction, get a clearer picture of the suspect's appearance and characteristics to speed up the process of crime solving. In summary, image superresolution reconstruction has broad applications in urban IoT systems.

Single image superresolution (SISR) is the task of generating high-resolution images using a single low-resolution image [32]. SISR algorithms are divided into three main categories: interpolation-based methods [33], reconstruction-based methods [34, 35], and learning-based methods [36,

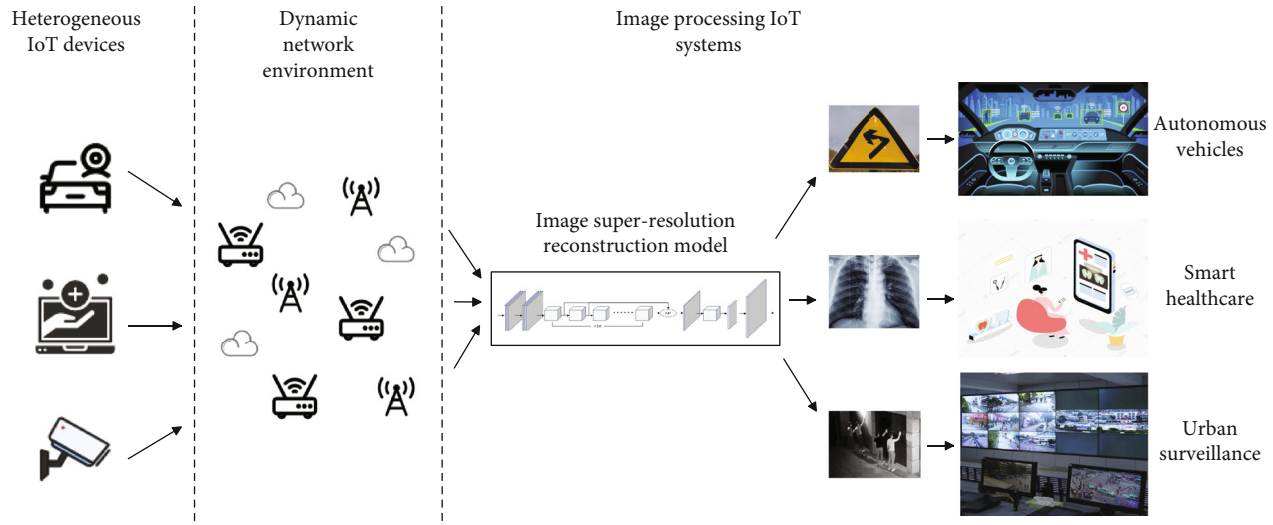

Figure 1: Image superresolution reconstruction for image processing IoT systems in Smart City.

37]. Learning-based methods are one of the most widely used methods at present. In particular, with the development of deep learning and generative adversarial networks, image superresolution has made great progress.

Dong et al. [38] proposed an SRCNN method that has realized end-to-end super resolution image reconstruction and better performance compared with other previous methods. However, the simple network structure limits its ability to extract features, and the MSE loss used by SRCNN stresses improving the image objective index, ignoring the subjective effect of the image. The detailed features of the blurred reconstructed images are VDSRs—depth models based on residual learning, which were proposed by Kim et al. [39]—that improve the model performance by introducing a residual structure, but there are problems such as large number of training parameters and unclear background of reconstructed images. EDSR proposed by Lim et al. [40] removes the BN layer and superimposes more layers to improve the reconstructed image quality by reducing the memory consumption of the BN layer. However, since L1 loss is used for training, the objective index of the reconstructed image is low.

Thanks to the generative adversarial networks proposed by Goodfellow et al. [41], the image superresolution task has opened a new chapter dominated by generative adversarial structures. SRGAN proposed by Ledig et al. [42] uses generative adversarial networks for image superresolution while using perceptual loss and adversarial loss to improve the realism of the reconstructed image, which makes the reconstructed image and the ground truth closer in semantics and style. However, the reconstructed image loses some high-frequency information due to the mere use of MSE loss to train the generator. ESRGAN proposed by Wang et al. [43] removes the BN layer based on SRGAN and introduces dense connections to avoid artifacts. VGG features before activation are used to improve perceptual loss and to make the edges and details of the reconstructed images clearer. The idea of relativistic GAN [44] is applied for reference to judge the probability that real images are more realistic than generated images in the discriminator, greatly enhancing the subjective effect of reconstructed images. Nevertheless, ESRGAN has many parameters and a long training time. RFB-ESRGAN proposed by Shang et al. [45] introduces a multi-scale receptive field module to extract edge features of images and alternately uses nearest-neighbor interpolation [46] and pixel-shuffle [47] in the upsampling module to promote the information interaction between network space and depth. However, asymmetric convolution in the multi-scale module can reduce the parameters and affect the accuracy of feature extraction, which is not conducive to restoring the original image's detailed features.

Due to the good performance of attention mechanisms in computer vision tasks represented by image classification [48], object detection [49], and semantic segmentation [50], Zhang et al. [51] first introduced channel attention into the image superresolution reconstruction task and proposed RCAN, which highlights the foreground information of reconstructed images to some extent. The SAN proposed by Dai et al. [52] uses a second-order attention network to capture distant spatial features, leveraging the underlying image features, and the reconstructed image color is closer to the original image. Liu et al. [53] proposed RFANet based on EDSR's proposed RFA module to exploit shallow residual features to achieve a good balance between model performance and parameter number and proposed an ESA spatial domain attention module to extract spatial domain features using stride length convolution and pooling instead of dilation convolution for dimensionality reduction to avoid the lack of image detail information caused by dilated convolution and achieve better results.

Although the above methods have achieved good results in image superresolution tasks, there remain problems such as the image foreground not being highlighted, lack of visual focus, etc. In this paper, we innovatively propose the parallel attention module (PAM) and use it as the basis to introduce skip connection and group convolution to build PAMNet, aiming to design a high-performance, high-quality image superresolution model that attends more to image

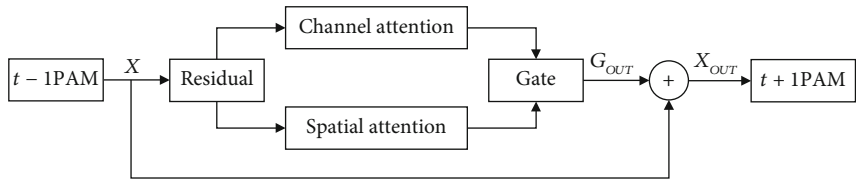

Figure 2: The structure of PAM in time t.

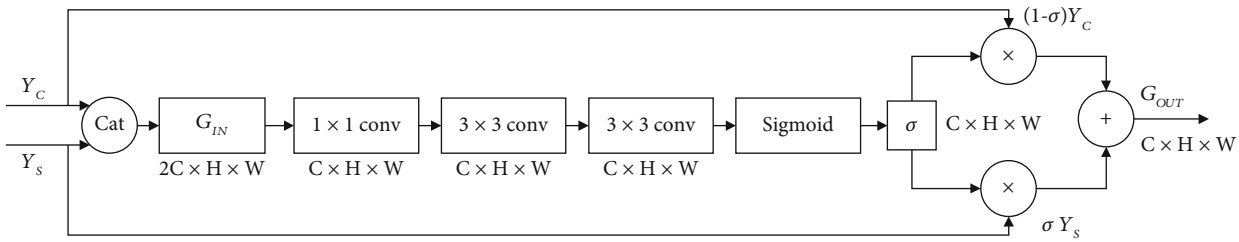

Figure 3: Gate network module.

foreground information and detailed features and has a smaller number of training parameters. The main contributions of this paper are as follows:

(1) Proposing a generic module named PAM, which computes channel attention and spatial attention in parallel on the residual block's residual branch, and then dynamically adjusts the weights of both using gated networks and nonuniform joint loss, so that the PAM module focuses on the attention domain with higher weights and thus can extract foreground information deeply

(2) Based on the PAM modules, we proposed PAMNet. By concatenating multiple PAM modules in PAMNet and introducing skip connections, the residual features from all the preceding PAM modules are fed directly to the PAM module at the end of the network for aggregation to leverage the shallow residual features, and the reconstructed images are closer to ground truth. In addition, by using group convolution, PAMNet is more lightweight than other methods

The reminder of the paper is organized as follows. Section 2 describes the PAM module, PAMNet, and the loss function of this paper in detail. Section 3 verifies the effectiveness and generality of this paper's method through ablation experiments and comparison experiments. Finally, Section 4 presents the conclusion of this study.

2. Method

2.1. PAM. The PAM module proposed in this paper can directly replace the residual block in the ResNet [54] backbone network, compute channel and spatial attention in parallel, splice the results in the channel dimension, and feed them into the gated network to extract the weight coefficients. In the backpropagation process, the channel attention and spatial attention weights are dynamically adjusted by nonuniform joint loss, focusing on extracting image foreground information in the attention domain with higher weights. The specific structure of PAM is shown in Figure 2.

In computing channel attention, a structure similar to SENet [55] is used, and the fully connected layer in SENet is replaced by 1×1 convolution, which can preserve the image's spatial features. The specific computation of channel attention is given by Eq. (1):

$$YC = X + CA(XR), \tag{1}$$

where $X \in R^{C \times H \times W}$ represents the input of the residual block, $XR \in R^{C \times H \times W}$ represents the output after computing the residuals, CA represents computing channel attention, and YC represents the final output of the channel attention. Meanwhile, in this paper, C represents the channel dimension of the feature map, H represents the height of the feature map, and W represents the width of the feature map, so that the three dimensions of a feature map can be represented as (C, H, W).

Referring to the HDC idea [56], PAM computes spatial attention using a three-layer cascaded dilation convolution with dilation rates of 1, 2, and 3. First, we use a 1×1 convolution to downscale the feature map with input dimensions (C, H, W) into a feature map with $(C/K, H, W)$ dimensions, where K is the downscaling factor, and in this paper, we take $K = 4$. Second, the feature map after downscaling is convoluted with three different expansion rates to expand the perceptual field with the minimum number of parameters in a finite number of steps to ensure the continuity of the perceptual domain and avoid the information loss caused by pooling. Finally, we use a 1×1 convolution to fuse the information of different channels of the feature map and go through Sigmoid activation to get the feature map weights in the $(1, H, W)$ dimension and assign the weights in the (H, W) dimension to multiply to the input feature map to focus on the image foreground information. The

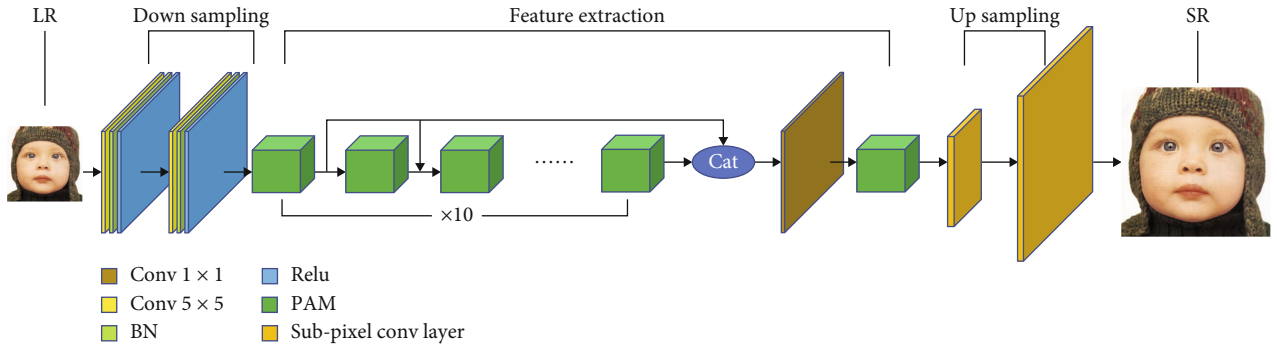

Figure 4: The structure of PAMNet.

specific computation of spatial attention is given by Eq. (2):

$$YS = X + SA(XR), \tag{2}$$

where $X \in R^{C\times H\times W}$ represents the input of the residual block, $XR \in R^{C\times H\times W}$ represents the output after computing the residuals, SA represents computing spatial attention, and YS represents the final output of spatial attention.

After obtaining the channel attention YC and spatial attention YS using the above method, the two are spliced in the channel dimension to obtain the input of the gated network$GIN \in R^{2C\times H\times W}$. Then, we use a 1×1 convolution to fuse the information and reduce the GIN dimension of (C, H, W). Then, two 3×3 convolutions for feature extraction and Sigmoid activation are used to obtain an activation output $\sigma \in R^{C\times H\times W}$ with values in the range (0, 1). Finally, the final output $GOUT \in R^{C\times H\times W}$ is obtained by multiplying σ by YC and YS as a linear combination of coefficients. Meanwhile, this weight is continuously updated during the backpropagation process, and the weights of channel attention and spatial attention are dynamically assigned in learning progress, focusing on extracting image foreground information in the attention domain with higher weights. The computation is given by Eq. (3):

$$\mathrm{GOUT} = (1-\sigma)YC + \sigma YS. \tag{3}$$

The specific structure of the gated network module is shown in Figure 3.

2.2. PAMNet. PAMNet is built with the PAM module as the core unit and postupsampling as the base structure, using skip connection, group convolution, and feature fusion. The network comprises a down-sampling layer, a feature extraction layer, and an upsampling layer. In this case, the downsampling layer uses a serial 5×5 convolution to initially extract image color, contour, and texture features. The upsampling layer uses pixel shuffle to enlarge the image. PAMNet benefits from the PAM module and skips connection, which focuses more on image foreground information reconstruction and can leverage shallow residual features, highlighting the visual focus of reconstructed images. The specific structure of PAMNet is shown in Figure 4.

The downsampling layer initially extracts the image's underlying features by two times 5×5 convolution and increases the number of feature map channels. The feature extraction layer and the upsampling layer are the core of PAMNet. The feature extraction layer uses PAM as the basic unit and serially multiple PAM modules to extract detailed features. The basic structure of the traditional residual block is two 3×3 same convolutions; serializing multiple blocks induces many parameters and complex computations, which seriously slows down the model's training. Therefore, in this paper, we use group convolution in the PAMNet feature extraction layer to reduce the number of parameters and add 1×1 convolution to fuse the group information. Taking the input feature map $XIN \in R^{C\times H\times W}$, output feature map $XOUT \in R^{C\times H\times W}$, and convolution kernel $F \in R^{C\times 3\times 3}$ as an example, the number of parameters of a residual block is given by Eq. (4):

$$PN = 3\times 3\times C\times C\times 2. \tag{4}$$

While using group convolution with group number g and 1×1 convolution, the number of parameters is reduced to Eq. (5):

$$\begin{aligned} PG &= \left(3\times 3\times C\times C\times \frac{1}{g} + 1\times 1\times C\times C\right)\times 2 \\ &= 2\times C\times C\times \left(3\times 3\times \frac{1}{g} + \right)1. \end{aligned} \tag{5}$$

According to the reference [57], take $C = 64$ and $g = 16$, we can get $PN = 73728$ and $PG = 12800$; we can see that the number of parameters using grouped convolution is only 17.36% of the normal convolution, which simplifies the number of parameters of the model while significantly increasing the training speed. Since the shallow residual features must pass through multiple computations before reaching the last PAM module, the deeper layers of the network fail to leverage the shallow information and lose some of the image's shallow features, which is inconvenient for reconstructing the image's color and texture information and severely limits the model's image reconstruction capability. Existing SR methods, such as RFANet, only use skip connection inside the RFA module, which fails to preserve

Table 1: Network training parameters.

Parameters	Values
Scale	4
Batch size	16
Optimizer	Adam
Learning rate	0.0002
PAM channels	64
Group of convolution	16

the shallow features of the image completely. In this study, we introduce a skip connection within PAMNet, and by skip connection, we input the residual features of all preceding PAM modules to the last PAM module in the feature extraction layer, reducing dimensionality and aggregating shallow features by 1×1 convolution. Compared with the simple stacking of multiple residual blocks, PAMNet retains the underlying image information so that it can participate in the subsequent computation to further extract the high-level semantic information while sending it directly to the end PAM module without any interference, which retains the underlying features and focuses on extracting the high-level image information.

The upsampling layer acts as the final layer of the network and is responsible for scaling the image to a specified magnification. Commonly used upsampling methods include linear interpolation, deconvolution [58], transposed convolution [59], subpixel convolution [60], and metaupscale [61]. Interpolation methods are the fastest, but reconstructed images are blurred and have low definition. Deconvolution and transposed convolution reconstruct images with a field of perception up to the same magnification as the image, which is not conducive to obtaining global features, and the reconstructed images are prone to checkerboard artifacts. Subpixel convolution has a larger field of perception and more contextual information, and the reconstructed image is clear in detail. The metaupscale does not need to determine the scale factor in advance, the image can be continuously enlarged by any factor, and the reconstructed image is high definition, which is often used for video superresolution reconstruction. Due to the faster computation speed of subpixel convolution and the high quality of reconstructed images, pixel shuffle is used for upsampling in this paper.

2.3. Loss Function. Similar to existing methods [42, 43, 51–53], this paper trains the network model based on the generative adversarial structure and optimizes the model parameters by the joint discriminator loss and generator loss, where discriminator loss LD is defined as Eq. (6):

$$LD = -\mathbb{E}xr[\log(D(xr, xf))] - \mathbb{E}xf[\log(1 - D(xf, xr)], \quad (6)$$

where xr is the real image, xf is the reconstructed image, and $D(xr, xf)$ computes the difference between the real image and the reconstructed image and uses the Sigmoid restriction $D(xr, xf) \in (0, 1)$.

Unlike the above methods, the generator loss in this paper comprises nonuniform joint loss, adversarial loss, and content loss. By using nonuniform joint loss, constraint the network learn image color and texture features while extracting more discriminative features and detailed information, focusing more on the reconstruction of image foreground information.

The nonuniform joint loss LU is based on L1 loss, and the reconstructed image and the original image are fed into the pretrained VGG-19 network to compute L1 loss LVGG1 before the first pooling layer and L1 loss LVGG2 before the last pooling layer by adjusting the weights of LVGG1 and L VGG2 to constrain the generator to extract the underlying features while learning more detailed information and discriminative features. The specific computation is given by Eq. (7):

$$LU = \alpha L\text{VGG1} + \beta\text{LVGG2}, \quad (7)$$

where α is the weight of LVGG1, and β is the weight of L VGG2; in this paper, we take $\alpha = 0.2$ and $\beta = 1$.

The adversarial loss LG is computed as in [11], and the specific computation is defined by Eq. (8):

$$LG = -\mathbb{E}xr[\log(1 - D(xr, xf))] - \mathbb{E}xf[\log(D(xf, xr)]. \quad (8)$$

Content loss LC computes the pixel difference between the real image and the reconstructed image using both L1 loss and L2 loss. Methods such as RFANet only use the L1 loss to compute the content loss, which induces the loss of some high-frequency information in the reconstructed images, and L1 loss is prone to sparse solutions and cannot be derived at the zero point, increasing the instability of GAN training. SRGAN only uses L2 loss to compute the content loss, which is influenced by outlier points. Although the reconstructed image has a higher peak signal-to-noise ratio (PSNR (dB)) but is prone to artifacts, the visual effect is poor, which opposes the original intention of image superresolution. PAMNet computes content loss using both L1 loss and L2 loss to enhance the method's robustness while reducing sparse solutions. The specific computation is given by Eq. (9):

$$LC = \mu L1(xr, xf) + \theta L2(xr, xf), \quad (9)$$

where xr represents the ground truth, xf represents the reconstructed image, and μ and θ represent the weight of L1 loss and L2 loss, respectively. In this study, we take $\mu = 0.75$ and $\theta = 0.25$.

In summary, the generator loss is defined by Eq. (10):

$$L = \gamma LG + \lambda LU + \eta LC, \quad (10)$$

where γ, λ, and η represent the weights of adversarial loss, nonuniform joint loss, and content loss. In this paper, we take $\gamma = 0.005$, $\lambda = 1$, and $\eta = 0.1$.

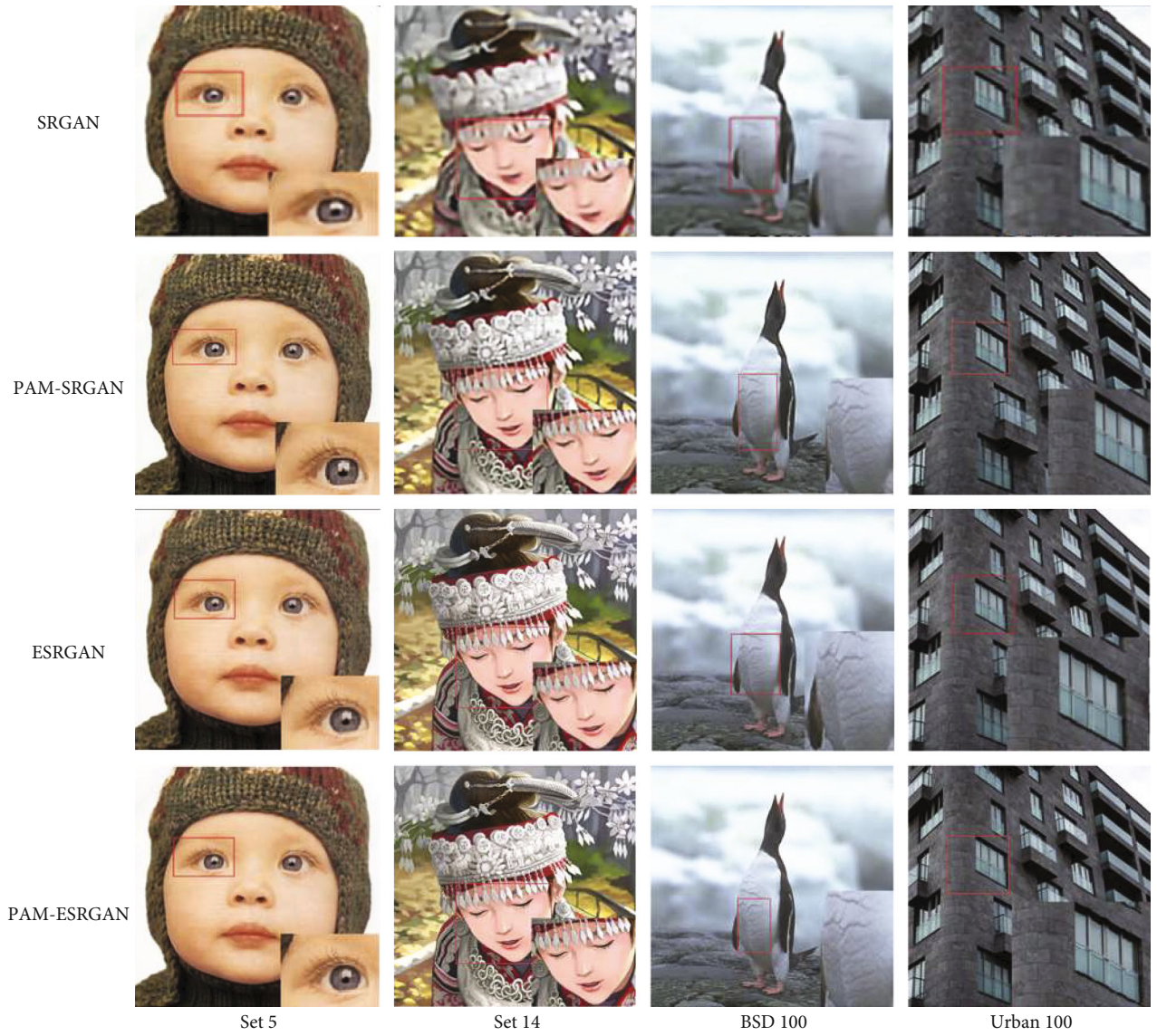

Figure 5: SRGAN, PAM-SRGAN, ESRGAN, and PAM-ESRGAN.

3. Experiment

3.1. Settings. Referring to the existing methods [40, 43, 51–53], to verify the effectiveness of this paper, we select 3450 images from DIV2K [62] and Flickr2K [63] as the training dataset and randomly select 60,000 subimages as the training images after cropping and mirror reversal operations on the original images. Meanwhile, we select Set5 [64], Set14 [65], BSD100 [66], and Urban100 [67] as the test datasets. The main parameters of the network are shown in Table 1.

This paper is based on PyTorch for experiments with the following hardware parameters: Intel i7 9700, NVIDIA 2080ti, and 32gRAM.

3.2. Results. In this paper, we focused on SISR reconstruction on a four-time deflation factor and used Set5, Set14, BSD100, and Urban100 as the test sets to compare with existing image superresolution methods from both subjective and objective aspects. We also embedded the PAM module into the backbone networks of SRGAN and ESRGAN to verify the effectiveness and generality of the module. Meanwhile, PSNR and SSIM were used as objective indices to quantify the quality of the reconstructed images.

3.2.1. Effectiveness and Generality of PAM. This section verifies the effectiveness and generality of the PAM module by replacing the basic residual block of SRGAN and the RRDB structure in ESRGAN using the PAM module and keeping the other structures and loss functions in the original network unchanged. The replaced models are called PAM-SRGAN and PAM-ESRGAN, and we selected the images in Set5, Set14, BSD100, and Urban100 for analysis. The results are shown in Figure 5.

The performance of SRGAN and ESRGAN with embedded PAM modules on different datasets is shown in Table 2.

Table 2: SRGAN, PAM-SRGAN, ESRGAN, and PAM-ESRGAN.

Method	Set5 PSNR/SSIM	Set14 PSNR/SSIM	BSD100 PSNR/SSIM	Urban100 PSNR/SSIM
SRGAN	29.40/0.847	26.02/0.739	25.16/0.669	24.29/0.661
PAM-SRGAN	31.24/0.883	27.93/0.749	25.91/0.701	24.83/0.692
ESRGAN	32.60/0.901	28.88/0.791	27.76/0.745	26.73/0.815
PAM-ESRGAN	32.74/0.902	28.91/0.787	27.80/0.745	26.94/0.813

Table 3: Numbers of PAM.

N	Set5 PSNR	Set14 PSNR	BSD100 PSNR	Urban100 PSNR
3	22.80	20.15	19.97	19.76
5	28.14	24.86	23.93	23.54
7	29.98	26.93	25.26	25.31
9	32.13	28.46	27.48	26.47
11	32.73	28.93	27.81	26.93

As seen in Table 2, PAM-SRGAN improves PSNR by 1.84 dB, 1.91 dB, 0.75 dB, and 0.54 dB over SRGAN on the four test sets; PAM-ESRGAN improves PSNR by 0.14 dB, 0.03 dB, 0.04 dB, and 0.21 dB over ESRGAN on the four test sets. The results show that the PAM module improves the performance of SRGAN and ESRGAN networks with good generality.

3.2.2. Performance of PAMNet. To give PAMNet the best performance, we performed the following experiments on the number of PAM modules in the feature extraction layer. Let the total number of PAM modules in PAMNet be N and $N \in [3, 5, 7, 9, 11]$. Keeping the other structures in PAMNet unchanged, the test results on different datasets are shown in Table 3.

As seen in Table 3, the performance of PAMNet outperforms SOTA method RFB-ESRGAN (32.66 dB, 28.88 dB, 27.79 dB, and 26.92 dB) and RFANet (32.72 dB, 28.91 dB, 27.77 dB, and 26.89 dB) when $N = 11$.

As the number of PAM modules (N) increases, the PSNR of PAMNet reconstructed images on different datasets grows accordingly. The variation relationship is shown in Figure 6 for the Urban100 dataset, for example.

Figure 6 shows that the PSNR of the reconstructed images does not continue to improve significantly with the increase in the number of PAM modules (N), and for a good balance between model performance and complexity, PAMNet takes $N = 11$. After determining the number of PAM modules and selecting images from the Set5, Set14, BSD100, and Urban100 test sets, one image from each test set was taken for analysis, and the results are shown in Figure 7.

As can be seen in Figure 7, due to the addition of a gated network and nonuniform joint loss in PAMNet, our method can produce sharper foreground information than existing methods (Figures 7(a) and 7(b)), and the detailed texture features of the reconstructed images are closer to Ground Truth (Figures 7(c) and 7(d)). In addition, PAMNet basically preserves the color and texture features of the image by introducing skip connection, and the overall image sharpness is basically on par with SOTA methods such as RFB-ESRGAN and RFANet.

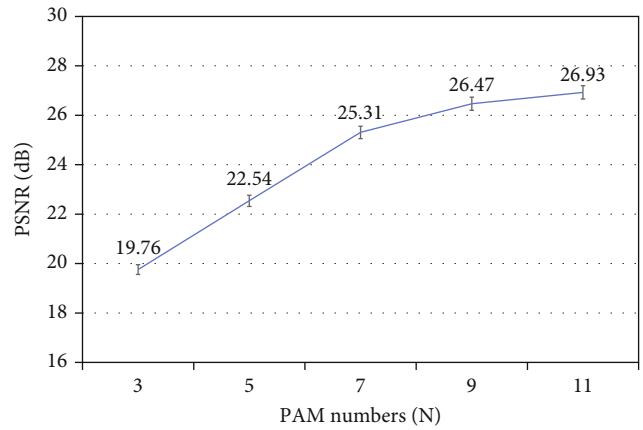

Figure 6: PAM numbers vs. PSNR on Urban100.

To verify the effectiveness of PAMNet from an objective perspective, we selected PSNR and SSIM as objective indices. The PSNR and SSIM of each image in Figure 7 are shown in Table 4.

As shown in Table 4, the PSNR and SSIM of PAMNet reconstructed images outperformed other methods, and only the RFB-ESRGAN method had slightly higher PSNR than PAMNet on Figure 6(a). To verify the generalization performance of PAMNet, the PSNR and SSIM of different methods on different test sets are shown in Table 5.

As shown in Table 5, the PSNR of PAMNet reconstructed images outperformed other methods in each test set, improving 0.01 dB, 0.02 dB, 0.04 dB, and 0.04 dB over RFB-ESRGAN in four test sets and improving 0.07 dB, 0.05 dB, 0.02 dB, and 0.01 dB over RFANet, and SSIM was Set5, and Urban100 datasets were slightly lower than RFB-ESRGAN. The experimental results show that, thanks to the PAM module and nonuniform joint loss, PAMNet can effectively extract image foreground information, improve the PSNR and SSIM of the reconstructed images, and enhance the foreground clarity while ensuring a clear background in the reconstructed images.

3.2.3. The Effect of Skip Connection. In this section, the skip connection in PAMNet was removed, and the other structures and loss functions were kept unchanged to investigate the effect of skip connection on PAMNet. The experimental results are presented in Table 6.

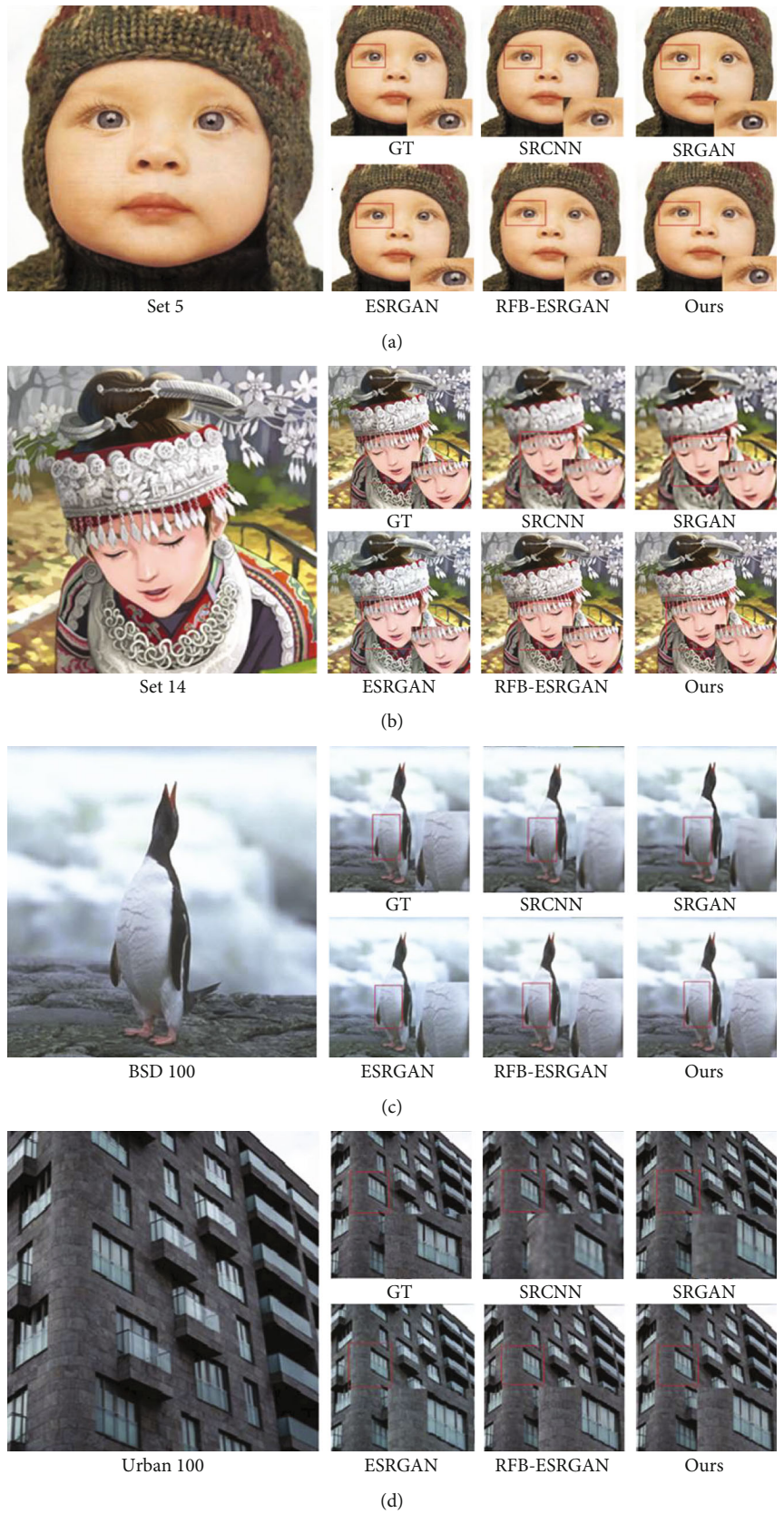

Figure 7: Reconstructed image of PAMNet: (a) from Set5, (b) from Set14, (c) from BSD100, and (d) from Urban100.

Table 4: Objective indicator comparison from Figure 7.

Method	Figure 7(a) PSNR/SSIM	Figure 7(b) PSNR/SSIM	Figure 7(c) PSNR/SSIM	Figure 7(d) PSNR/SSIM
SRCNN	29.04/0.815	29.28/0.786	26.72/0.748	26.95/0.782
SRGAN	30.11/0.862	28.06/0.775	27.02/0.743	27.29/0.779
ESRGAN	33.87/0.911	29.34/0.825	28.03/0.796	27.93/0.813
RFB-ESRGAN	33.91/0.912	31.02/0.843	28.13/0.797	28.07/0.823
PAMNet (ours)	33.89/0.916	32.65/0.851	28.41/0.804	28.24/0.830

Table 5: Average objective indicator.

Method	Set5 PSNR/SSIM	Set14 PSNR/SSIM	BSD100 PSNR/SSIM	Urban100 PSNR/SSIM
SRCNN	30.48/0.862	27.50/0.751	26.90/0.710	24.52/0.722
SRGAN	29.40/0.847	26.02/0.739	25.16/0.669	24.29/0.661
VDSR	31.35/0.883	28.02/0.768	27.29/0.7260	25.18/0.754
EDSR	32.46/0.896	28.80/0.787	27.71/0.742	26.64/0.803
DBPN	32.47/0.898	28.82/0.786	27.72/0.740	26.38/0.794
ESRGAN	32.60/0.901	28.88/0.791	27.76/0.745	26.73/0.815
SAN	32.64/0.900	28.92/0.788	27.78/0.743	26.79/0.806
RFB-ESRGAN	32.72/0.902	28.91/0.801	27.77/0.746	26.89/0.817
RFANet	32.66/0.900	28.88/0.789	27.79/0.744	26.92/0.811
PAMNet(ours)	32.73/0.901	28.93/0.802	27.81/0.750	26.93/0.816

Table 6: Skip connection in PAMNet.

Name	Set5 PSNR/SSIM	Set14 PSNR/SSIM	BSD100 PSNR/SSIM	Urban100 PSNR/SSIM
No skip connection	32.21/0.894	28.36/0.801	27.65/0.744	26.69/0.812
Skip connection	32.73/0.901	28.930.802	27.81/0.750	26.93/0.816

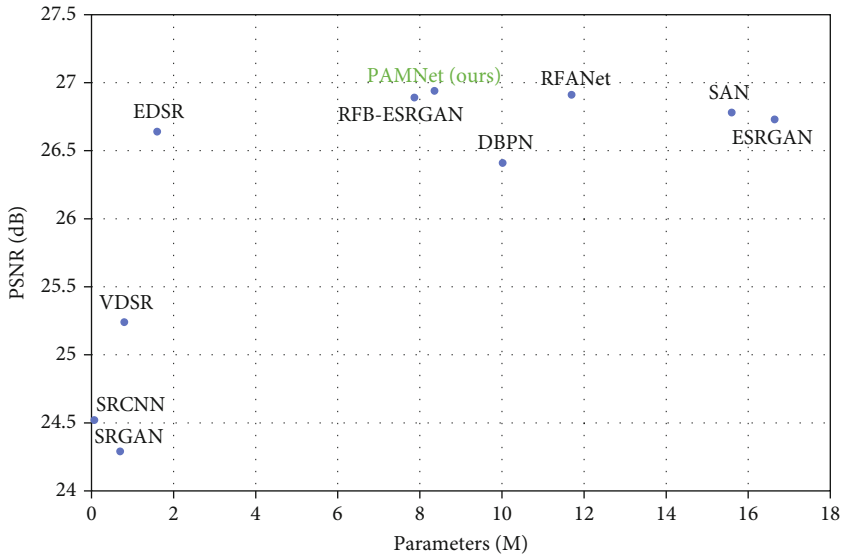

Figure 8: PSNR vs. parameters on Urban100 (×4).

As shown in Table 6, the PSNR of PAMNet reconstructed images on different test sets decreased by 0.52 dB, 0.57 dB, 0.16 dB, and 0.24 dB, after removing the skip connection in PAMNet, and the performance of PAMNet decreased significantly, which constrained the utilization of shallow features by the model. The experimental results show that the skip connection significantly impacted PAMNet, and the use of skip connection could improve the utilization of shallow features in PAMNet, thereby enhancing the comprehensive performance of the model.

3.2.4. Model Complexity. To evaluate the complexity of the PAMNet model, it was compared with existing SR methods: SRCNN, SRGAN, VDSR, EDSR, DBPN, SAN, ESRGAN, RFB-ESRGAN, and RFANet. The results are shown in Figure 8.

As seen in Figure 8, PAMNet has smaller parameters and better performance than DBPN, RFANet, SAN, and ESRGAN. Compared with RFB-ESRGAN, PAMNet has a slightly larger number of parameters but slightly outperforms RFB-ESRGAN overall.

4. Conclusion

In this paper, we proposed a generic PAM module for image superresolution reconstruction to extract foreground information and high-frequency features of images. The module computed channel attention and spatial attention in parallel and used the gated network to extract the two-weight coefficients and cooperated with the nonuniform joint loss to dynamically modify the two weights during the backpropagation process, so that the network attended more to the extraction of foreground information and discriminative features. To fully reflect the good performance of PAM modules, PAMNet was further proposed to connect multiple PAM modules in series in PAMNet. The ablation experiments verified the effectiveness and generality of the PAM module and the necessity of skip connection in PAMNet. By contrast experiments with existing state-of-the-art image superresolution methods, the average PSNR improvement of PAMNet on different data sets is 0.4 dB, and the average SSIM improvement is 0.005. It is verified that PAMNet achieves a good balance between performance and model complexity. By using PAMNet, in many applications of urban IoT systems, such as autonomous vehicles, smart healthcare, and urban surveillance, it is possible to generate clearer and more foreground-focused high-resolution images than existing image superresolution methods, improving the reliability of urban IoT systems and satisfying people's vision of a better life. Limited by the training equipment and the research content, only the image superresolution method on the ×4 magnification factor has been studied. In the future, we will also continue to research faster and greater magnification image superresolution methods, so that various smart technologies can continue to benefit humanity and all families.

Data Availability

The open datasets used to support the findings of this study are included within the article. The link is as follows: https://data.vision.ee.ethz.ch/cvl/DIV2K/.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is partially supported by Telecommunications Advancement Foundation (Japan) Research Grant, RIEC Nationwide Cooperative Research Projects, Research Institute of Electrical Communication, Tohoku University, Japan, H31/B18, and ROIS NII Open Collaborative Research 2021 (21FA03).

References

[1] K. Yu, Z. Guo, Y. Shen, J. C. W. Lin, T. Sato, and T. Sato, "Secure artificial intelligence of things for implicit group recommendations," *IEEE Internet of Things Journal*, 2021.

[2] L. Tan, K. Yu, F. Ming, X. Cheng, and G. Srivastava, "Secure and resilient artificial intelligence of things: a HoneyNet approach for threat detection and situational awareness," *IEEE Consumer Electronics Magazine*, p. 1, 2021.

[3] C. Feng, K. Yu, M. Aloqaily, M. Alazab, Z. Lv, and S. Mumtaz, "Attribute-based encryption with parallel outsourced decryption for edge intelligent IoV," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 13784–13795, 2020.

[4] Z. Guo, K. Yu, A. Jolfaei, A. K. Bashir, A. O. Almagrabi, and N. Kumar, "fuzzy detection system for rumors through explainable adaptive learning," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 12, pp. 3650–3664, 2021.

[5] F. Ding, G. Zhu, Y. Li, X. Zhang, P. K. Atrey, and S. Lyu, "Anti-Forensics for face swapping videos via adversarial training," *IEEE Transactions on Multimedia*, p. 1, 2021.

[6] F. Ding, G. Zhu, M. Alazab, X. Li, and K. Yu, "Deep-learning-empowered digital forensics for edge consumer electronics in 5G HetNets," *IEEE Consumer Electronics Magazine*, p. 1, 2020.

[7] L. Tan, K. Yu, L. Lin et al., "Speech emotion recognition enhanced traffic efficiency solution for autonomous vehicles in a 5G-enabled space-air-ground integrated intelligent transportation system," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–13, 2021.

[8] J. Zhang, X. Hu, Z. Ning et al., "Energy-latency tradeoff for energy-aware offloading in mobile edge computing networks," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2633–2645, 2017.

[9] Z. Ning, X. Hu, Z. Chen et al., "A cooperative quality-aware service access system for social internet of vehicles," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2506–2517, 2017.

[10] B. Hu, G. P. Gao, L. L. He, X. D. Cong, and J. N. Zhao, "Bending and on-arm effects on a wearable antenna for 2.45 GHz body area network," *IEEE Antennas and Wireless Propagation Letters*, vol. 15, pp. 378–381, 2016.

[11] X. Hu, J. Cheng, M. Zhou et al., "Emotion-aware cognitive system in multi-channel cognitive radio ad hoc networks," *IEEE Communications Magazine*, vol. 56, no. 4, pp. 180–187, 2018.

[12] K. Yu, L. Tan, S. Mumtaz et al., "Securing critical infrastructures: Deep-Learning-Based threat detection in IIoT," *IEEE Communications Magazine*, vol. 59, no. 10, pp. 76–82, 2021.

[13] H. Li, K. Yu, B. Liu, C. Feng, Z. Qin, and G. Srivastava, "An efficient ciphertext-policy weighted attribute-based encryption for the internet of health things," *IEEE Journal of Biomedical and Health Informatics*, vol. PP, 2021.

[14] L. Zhen, Y. Zhang, K. Yu, N. Kumar, A. Barnawi, and Y. Xie, "Early collision detection for massive random access in satellite-based internet of things," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 5, pp. 5184–5189, 2021.

[15] Z. Guo, K. Yu, Y. Li, G. Srivastava, and J. C. W. Lin, "Deep learning-embedded social internet of things for ambiguity-aware social recommendations," *IEEE Transactions on Network Science and Engineering*, 2021.

[16] T. Guo, K. Yu, M. Aloqaily, and S. Wan, "Constructing a prior-dependent graph for data clustering and dimension reduction in the edge of AIoT," *Future Generation Computer Systems*, vol. 128, pp. 381–394, 2022.

[17] Y. Gong, L. Zhang, R. Liu, K. Yu, and G. Srivastava, "Nonlinear MIMO for industrial Internet of Things in cyber–physical systems," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5533–5541, 2021.

[18] K. Yu, M. Arifuzzaman, Z. Wen, D. Zhang, and T. Sato, "A key management scheme for secure communications of information centric advanced metering infrastructure in smart grid," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 8, pp. 2072–2085, 2015.

[19] C. Feng, B. Liu, Z. Guo, K. Yu, Z. Qin, and K. K. R. Choo, "Blockchain-based cross-domain authentication for intelligent 5G-enabled internet of drones," *IEEE Internet of Things Journal*, 2021.

[20] C. Feng, B. Liu, K. Yu, S. K. Goudos, and S. Wan, "Blockchain-empowered decentralized horizontal federated learning for 5G-enabled UAVs," *IEEE Transactions on Industrial Informatics*, p. 1, 2021.

[21] K. Yu, L. Tan, C. Yang et al., "A blockchain-based shamir's threshold cryptography scheme for data protection in industrial internet of things settings," *IEEE Internet of Things Journal*, 2021.

[22] F. Ding, K. Yu, Z. Gu, X. Li, and Y. Shi, "Perceptual enhancement for autonomous vehicles: restoring visually degraded images for context prediction via adversarial training," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–12, 2021.

[23] L. Zhao, H. Li, N. Lin, M. Lin, C. Fan, and J. Shi, "Intelligent content caching strategy in autonomous driving Toward 6G," *IEEE Transactions on Intelligent Transportation Systems (T-ITS)*, pp. 1–11, 2021.

[24] L. Zhao, W. Zhao, A. Hawbani et al., "Novel online sequential learning-based adaptive routing for edge software-defined vehicular networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 5, pp. 2991–3004, 2021.

[25] K. Yu, L. Tan, L. Lin, X. Cheng, Z. Yi, and T. Sato, "Deep-learning-empowered breast cancer auxiliary diagnosis for 5GB remote e-health," *IEEE Wireless Communications*, vol. 28, no. 3, pp. 54–61, 2021.

[26] L. Tan, K. Yu, N. Shi, C. Yang, W. Wei, and H. Lu, "Towards secure and privacy-preserving data sharing for COVID-19 medical records: a blockchain-empowered approach," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 1, pp. 271–281, 2021.

[27] Y. Sun, J. Liu, K. Yu, M. Alazab, and K. Lin, "PMRSS: privacy-preserving medical record searching scheme for intelligent diagnosis in IoT healthcare," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 1981–1990, 2022.

[28] W. Shang, J. Chen, H. Bi, Y. Sui, Y. Chen, and H. Yu, "Impacts of COVID-19 pandemic on user behaviors and environmental benefits of bike sharing: a big-data analysis," *Applied Energy*, vol. 285, no. 116429, p. 116429, 2021.

[29] H. Peng, B. Hu, Q. Shi et al., "Removal of ocular artifacts in EEG—an improved approach combining DWT and ANC for portable applications," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 3, pp. 600–607, 2013.

[30] L. Yang, K. Yu, S. X. Yang, C. Chakraborty, Y. Lu, and T. Guo, "An intelligent trust cloud management method for secure clustering in 5G enabled internet of medical things," *IEEE Transactions on Industrial Informatics*, p. 1, 2021.

[31] D. Wang, Y. He, K. Yu, G. Srivastava, L. Nie, and R. Zhang, "Delay sensitive secure NOMA transmission for hierarchical HAP-LAP medical-care IoT networks," *IEEE Transactions on Industrial Informatics*, p. 1, 2021.

[32] J. L. Harris, "Diffraction and resolving power," *Journal of the Optical Society of America*, vol. 54, no. 7, pp. 931–936, 1964.

[33] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super-resolution: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 10, pp. 3365–3387, 2020.

[34] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays, using convex projections," *Journal of the Optical Society of America A*, vol. 6, no. 11, pp. 1715–1726, 1989.

[35] M. Irani and S. Peleg, "Super resolution from image sequences," in *Proceedings. 10th International Conference on Pattern Recognition*, pp. 115–120, Atlantic City, USA, 1990.

[36] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp. 56–65, 2002.

[37] C. Ma, C. Y. Yang, X. Yang, and M. H. Yang, "Learning a no-reference quality metric for single-image super-resolution," *Computer Vision and Image Understanding*, vol. 158, pp. 1–16, 2017.

[38] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2016.

[39] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016*, pp. 1646–1654, Las Vegas, NA, USA, 2016.

[40] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2017, pp. 136–144, Hawaii, HI, USA, 2017.

[41] I. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial nets," *Advances in Neural Information Processing Systems*, vol. 27, pp. 2672–2680, 2014.

[42] C. Ledig, L. Theis, F. Huszár et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE Conference on Computer*

Vision and Pattern Recognition, CVPR 2017, pp. 4681–4690, Hawaii, HI, USA, 2017.

[43] X. Wang, K. Yu, S. Wu et al., "Esrgan: enhanced super-resolution generative adversarial networks," in *Proceedings of the European Conference on Computer Vision, ECCV 2018*, Munich, Germany, 2018.

[44] A. Jolicoeur-Martineau, "The relativistic discriminator: a key element missing from standard GAN," in *International Conference on Learning Representations, ICLR 2019*, New Orleans, Louisiana, USA, 2019.

[45] T. Shang, Q. Dai, S. Zhu, T. Yang, and Y. Guo, "Perceptual extreme super-resolution network with receptive field block," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPR 2020*, pp. 440-441, Seattle, WA, USA, 2020.

[46] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, ""When is "nearest neighbor" meaningful?"," in *International Conference on Database Theory*, pp. 217–235, Berlin, Heidelberg, 1999.

[47] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition, CVPR 2018*, pp. 2472–2481, Salt Lake City, UT, USA, 2018.

[48] T. Durand, N. Mehrasa, and G. Mori, "Learning a deep Convnet for multi-label classification with partial labels," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2019*, pp. 647–657, Long Beach, CA, USA, 2019.

[49] T. Wang, T. Yang, and M. Danelljan, "Learning human-object interaction detection using interaction points," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020*, pp. 4116–4125, Seattle, WA, USA, 2020.

[50] C. Yu, J. Wang, C. Gao, G. Yu, C. Shen, and N. Sang, "Context prior for scene segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020*, pp. 12416–12425, Seattle, WA, USA, 2020.

[51] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European Conference on Computer Vision, ECCV 2018*, pp. 286–301, Munich, Germany, 2018.

[52] T. Dai, J. Cai, Y. Zhang, S. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2019*, pp. 11065–11074, Long Beach, CA, USA, 2019.

[53] J. Liu, W. Zhang, Y. Tang, J. Tang, and G. Wu, "Residual feature aggregation network for image super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020*, pp. 2359–2368, Seattle, WA, USA, 2020.

[54] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016,*, pp. 770–778, Las Vegas, NV, USA, 2016.

[55] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018*, pp. 7132–7141, Salt Lake City, UT, USA, 2018.

[56] Y. Yuan, K. Yang, and C. Zhang, "Hard-aware deeply cascaded embedding," in *Proceedings of the IEEE International Conference on Computer Vision, ICCV 2017*, pp. 814–823, Venice, Italy, 2017.

[57] Y. Ioannou, D. Robertson, R. Cipolla, and A. Criminisi, "Deep roots: improving cnn efficiency with hierarchical filter groups," in *Proceedings of the IEEE conference on computer vision and pattern recognition, CVPR 2017*, pp. 1231–1240, Honolulu, HI, USA, 2017.

[58] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2010*, pp. 2528–2535, San Francisco, CA, USA, 2010.

[59] H. Gao, H. Yuan, Z. Wang, and S. Ji, "Pixel transposed convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 5, pp. 1218–1227, 2020.

[60] W. Shi, J. Caballero, F. Huszar et al., "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016*, pp. 1874–1883, Las Vegas, NV, USA, 2016.

[61] X. Hu, H. Mu, X. Zhang, Z. Wang, T. Tan, and J. Sun, "Meta-SR: a magnification-arbitrary network for super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2019*, pp. 1575–1584, Long Beach, CA, USA, 2019.

[62] E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: dataset and study,," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops, CVPRW2017*, pp. 126–135, Honolulu, HI, USA, 2017.

[63] R. Timofte, E. Agustsson, L. Van Gool, M. H. Yang, and L. Zhang, "Ntire 2017 challenge on single image super-resolution: methods and results," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops, CVPRW2017*, pp. 114–125, Honolulu, HI, USA, 2017.

[64] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proceedings of the 23rd British Machine Vision Conference, BMVC 2012*, Guildford, U.K., 2012.

[65] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *International conference on curves and surfaces*, pp. 711–730, Springer, Berlin, Heidelberg, 2010.

[66] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proceedings Eighth IEEE International Conference on Computer Vision, ICCV 2001*, vol. 2, pp. 416–423, Vancouver, Canada, 2001.

[67] J. B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proceedings of the IEEE conference on computer vision and pattern recognition, CVPR 2015*, pp. 5197–5206, Boston, MA, USA, 2015.